WHAT IS CLAIMED IS:

1. A method to identify text-like pixels from an image, the method comprising:

   (a) providing an image; and

   (b) classifying line segments of pixels within the image by edge-bounded averaging.

2. The method of claim 1, further comprising:

   (c) examining sub-blobs of pixels within the image; and

   (d) performing sub-blob connectivity analysis.

3. The method of claim 2, further comprising:

   (e) identifying and classifying edges of pixels within the image;

   (f) performing filling to further classify pixels within the image;

   (g) performing consistency analysis of pixels within the image;

   (h) performing pixel connectivity analysis of pixels within the image; and

   (i) identifying text pixels within the image.

4. The method of claim 1, wherein the image is a digital image.

5. The method of claim 1, further comprising performing color space conversion of the image.

6. The method of claim 1, further comprising smoothing the image.

7. The method of claim 1, wherein a Gaussian lowpass filter is applied to the image, the filter being $f_{i,j} = ke^{-\alpha^2\left[(i-c)^2+(j-c)^2\right]/c^2}$ where k is a normalizing factor such that $\sum_{i,j} f_{i,j} = 1.0$, c is the center of the filter and $\alpha = 1.0$.

8. The method of claim 3, step (e) identifying and classifying edges of pixels within the image, wherein every pixel is classified as NON EDGE, WHITE EDGE or BLACK EDGE.

9. The method of claim 8, wherein step (e) identifying and classifying edges of pixels within the image comprises:

(1) calculating a vertical gradient $G^I_{i,j}$, a horizontal gradient $G^J_{i,j}$ and the magnitude of gradient $M_{i,j}$ using the formula,

$$G^I_{i,j} = (y_{i+1,j-1} + 2y_{i+1,j} + y_{i+1,j+1}) - (y_{i-1,j-1} + 2y_{i-1,j} + y_{i-1,j+1})$$
$$G^J_{i,j} = (y_{i+1,j+1} + 2y_{i,j+1} + y_{i-1,j+1}) - (y_{i+1,j-1} + 2y_{i,j-1} + y_{i-1,j-1})$$
$$M_{i,j} = \sqrt{(G^I_{i,j})^2 + (G^J_{i,j})^2}$$

Where $y_{i,j}$ is a pixel luminance value at an index i,j.

(2) calculating a discrete Laplacian (a second directive):

$$L_{i,j} = (y_{i-2,j} + y_{i+2,j} + y_{i,j-2} + y_{i,j+2}) - 4y_{i,j}$$

(3) classifying every pixel as the following:

```
If $M_{i,j} > T_e$ then
    If $L_{i,j} < 0$
        Classify pixel at (i,j) as WHITE EDGE.
    Else
        Classify pixel at (i,j) as BLACK EDGE.
    Endif
Else
    Classify pixel at (i,j) as NON EDGE.
Endif
```

10. The method of claim 1, wherein step (b) classifying line segments of pixels within the image by edge-bounded averaging comprises:

starting from a first side of a line proceeding to a second side of the line identifying consecutive segments of pixels as NON EDGE, WHITE EDGE or BLACK EDGE.

11. The method of claim 1, wherein step (b) classifying line segments of pixels within the image by edge-bounded averaging comprises:

computing the edge-bounded averaging for at least eight locations including both end points of a central interior, both end points of a left edge segment, both end points of a right edge segment, a right end point of a left interior and a left end point of a right interior.

12. The method of claim 11, further comprising:

classifying the central interior as NON TEXT, BLACK INTERIOR or WHITE INTERIOR based upon the edge-bounded averaging values.

13. The method of claim 3, wherein step (f) performing filling to further classify pixels within the image comprises:

classifying segments as NON TEXT; and

examining segments classified as NON TEXT to determine whether they may be reclassified as BLACK INTERIOR, BLACK EDGE, WHITE INTERIOR or WHITE EDGE.

14. The method of claim 3, wherein step (g) performing vertical consistency analysis of pixels within the image comprises:

examining pixels not yet classified as NON TEXT to determine whether they are BLACK INTERIOR, BLACK EDGE, WHITE INTERIOR or WHITE EDGE.

15. The method of claim 3, wherein step (h) performing pixel connectivity analysis of pixels within the image comprises:

identifying aggregates of pixels having been identified as candidates for text, the aggregates being sub-blobs; and

collecting statistics with respect to each sub-blob, wherein said statistics are selected from the group consisting of total number of pixels, sums of color values, number of border pixels, number of broken border pixels and horizontal run length.

16. The method of claim 2, wherein step (c) examining sub-blobs of pixels within the image comprises:

examining each sub-blob to determine whether it is NON TEXT.

17. The method of claim 3, wherein step (i) identifying text pixels comprises:

examining each sub-blob to classify each pixel as either a text pixel or a non-text pixel.

18. A method to identify text-like pixels from an image, the method being directed to a compound document image compression scheme, the method comprising the steps of:

    (a) providing an image;

    (b) identifying and classifying edges of pixels within the image;

    (c) classifying line segments of pixels within the image by edge-bounded averaging;

    (d) performing vertical filling to further classify pixels within the image;

    (e) performing vertical consistency analysis of pixels within the image;

    (f) performing pixel connectivity analysis of pixels within the image; and

    (g) examining sub-blobs of pixels within the image.

19. The method of claim 18, further comprising:

    outputting a two layer image representation compatible with PDF Reference 1.2.

20. The method of claim 18, wherein step (e) performing pixel connectivity analysis of pixels within the image comprises:

identifying aggregates of pixels having been identified as candidates for text, the aggregates being sub-blobs;

collecting each sub-blobs statistics: total number of pixels, sums of color values, number of border pixels, number of broken border pixels and horizontal run length; and

counting sums of each luminance and chroma.

21. The method of claim 18, further comprising:

    outputting a three layer image representation compatible with PDF Reference 1.3.

22. The method of claim 18, wherein step (e) performing pixel connectivity analysis of pixels within the image comprises:

identifying aggregates of pixels having been identified as candidates for text, the aggregates being sub-blobs;

collecting each sub-blobs statistics: total number of pixels, sums of color values, number of border pixels, number of broken border pixels and horizontal run length; and

counting sums of each Y, C_r, C_b.

23. A system for identifying text-like pixels from an image, the system comprising:

a CPU running software adapted to:

(a) classify line segments of pixels within the image by edge-bounded averaging.

24. The system of claim 25, wherein the software is further adapted to:

(b) examine sub-blobs of pixels within the image; and

(c) perform sub-blob connectivity analysis.

25. The system of claim 26, wherein the software is further adapted to:

(d) identify and classify edges of pixels within the image;

(e) perform vertical filling to further classify pixels within the image;

(f) perform vertical consistency analysis of pixels within the image;

(g) perform pixel connectivity analysis of pixels within the image; and

(h) identify text pixels.